

Condensed matter theory by computer simulation: from materials to chemical biology

[DOI: 10.1051/EPN:2007009]

Mauro Ferrario¹, Giovanni Ciccotti² and Kurt Binder³,¹ Dipartimento di Fisica, Università di Modena e Reggio Emilia • Via Campi, 213/A, 41100 Modena • Italy² Dipartimento di Fisica, INFN, Università di Roma La Sapienza • Piazzale Aldo Moro 2, 00185 Roma • Italy³ Institut für Physik, Johannes Gutenberg-Universität • D-55099 Mainz • Germany

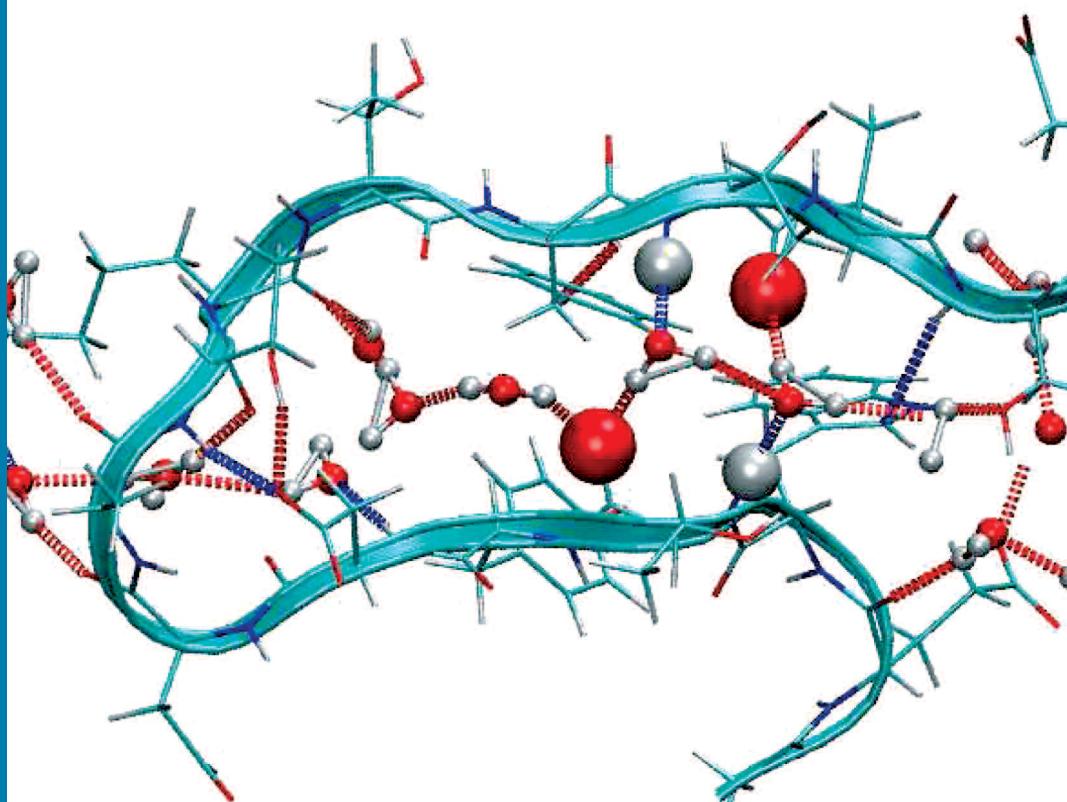
Condensed matter systems, ranging from simple fluids and solids to complex multi-component materials and even biological matter, are governed by well understood laws of physics: on the relevant scales of length and time; the appropriate description would be just the Schrödinger equation for the quantum-many-body problem of the nuclei and electrons interacting with Coulomb forces. Statistical mechanics would then provide the framework to extend this quantum theory of condensed matter towards a statistical description in terms of averages taken at nonzero temperature.

However, this program cannot be carried out straightforwardly: even dealing only with the associated electrons is still a challenge for the methods of quantum chemistry. Similarly, standard statistical mechanics makes precise explicit statements only on the properties of systems for which the many-body problem can be effectively reduced to a problem of independent particles or quasi-particles. Such problems are, for instance, ideal gases, paramagnets, or the multidimensional harmonic oscillator describing phonons in harmonic crystals. While all these problems are useful and educative (we hence teach them all to students to illustrate the spirit of the general theoretical framework) they do not encompass most of the

problems of interest in condensed matter physics: the interactions among the considered degrees of freedom introduce nontrivial correlations between them. Systematic perturbation-theory type methods usually do not lead very far, and uncontrolled closed form approximations often fail utterly. A well-known example is the study of the liquid gas transition in fluids: simple theories such as the van der Waals equation and its extensions (*e.g.* the "perturbed chain" statistical associating fluid theory [PC-SAFT] intended to model the equation of state of polymer melts [1]) give rise to spurious loops in the isotherms, fail to describe the critical behavior correctly, and may even predict completely unphysical phase equilibria that do not correspond to any physical behavior of the system but are mere artefacts of the inappropriate approximations [1]. The theory of condensed matter systems contains a never-ending list of such failures!

Hence, until about 50 years ago, condensed matter theory suffered from the basic problem that only a formal framework for the theoretical description in terms of quantum theory and statistical mechanics did exist, while a reliable set of tools that would allow accurate explicit predictions to be made for the static and dynamic properties of these systems from first principles was simply missing!

This unsatisfactory situation has changed fundamentally through the invention of computer simulation, which provides a much more promising approach to these problems. Computer simulation is a novel methodic route allowing the



◀ Fig. 1: Typical transition state configuration of the 16 residue terminal fragment of protein G-B1, that forms a so-called β -hairpin structure. The two strands of the backbone, which is represented as a ribbon, are separated by a strip of water molecules. In particular the important backbone hydrogen bonds 3-4 (indicated by large spheres) are bridged by water molecules. Reprinted with permission from [2].

treatment of strong correlations in many-body systems. Starting with the introduction of Monte Carlo (MC) methods more than 50 years ago, and Molecular Dynamics (MD) methods shortly thereafter, the scope of simulation methods has been rapidly expanding, and many exciting technical extensions are still under development, providing broad applications for this new theoretical paradigm of science, and new opportunities for important discoveries. While in the early days of computer simulation suitable computing facilities were very rare and, compared to present-day facilities, very slow and with very small storage capacities, very important discoveries were immediately made, which were very surprising at that time. These surprises include the fact that hard spheres crystallize at a density long before close packing has been achieved, and that dynamic correlations in fluids exhibit long time tails. These discoveries not only have greatly influenced the thinking of theorists and experimentalists alike, but have also been the starting point of a great variety of methodological developments. Now, when a powerful desktop computer is at the hand of every scientist, these advances in simulation methods progress at a breathtaking speed! These methods range from basic many-body quantum mechanics to classical and quantum-mechanical MC and MD methods, both in equilibrium and far from it, and to techniques on a coarse-grained mesoscopic level such as the so-called "Lattice Boltzmann" and "dissipative particle dynamics" methods. The fields to which these methods can be applied are extremely diverse: from fundamental problems in astrophysics and high energy particle physics (such as the transition from the quark gluon plasma to hadronic matter) to basic phenomena of solid state physics (such as magnetism, superconductivity and phase transitions between different crystal structures) and to the rich world of phenomena that complex fluids (liquid crystals, polymers, colloids, micro-emulsions, membranes, etc.) exhibit. Computer simulations in condensed matter now can elucidate properties of materials as well as the major nonequilibrium processes that take place in the living cell.

The school, held at the Ettore Majorana Foundation and Center for Scientific Culture (EMFCSC), Erice (Sicily), in July 2005 emphasized precisely these topics, and was properly entitled "*Computer Simulations in Condensed Matter: From Materials to Chemical Biology*". The precise selection of topics received special motivation by the research interests of Prof. Mike Klein (Univ. of Pennsylvania), who was honored at this event on the occasion of his 65th birthday. Although in the remainder of this article we cannot describe comprehensively the contents of the lecture notes of this school (which take about 1300 printed pages [2]), we hope nevertheless to sum-

marize concisely where is the front line of current research, using the chapters of these Lecture Notes [2] as a useful guide on this battleground.

Basic tools: classical and quantum simulations methods

As soon as the Metropolis importance sampling method (MC) and molecular dynamics (MD), first proposed in 1953 [3] and in 1957 [4], were introduced, they were applied to elucidate the equation of state of hard disks. Subsequently the study of phase transitions (both of lattice models and of off-lattice systems) has been a very important task of MC and MD methods, since analytical methods are rarely reliable for such problems [5-7]. However, a naive application of sampling techniques to such phenomena will also encounter some problems: at first order transitions, the states corresponding to the two competing phases are normally separated by huge free energy barriers in phase space [5-7]. Since the sampling algorithms do not yield any information on the partition function itself, the information on absolute free energies is lacking and an accurate sampling of the free energy difference between the states is mandatory. This requires (with any standard algorithm) the observation of a large number of transitions back and forth between the two states. Conversely, at second order transitions the growth of the order parameter correlation length (which ultimately diverges when one approaches criticality) and the associated slowness of relaxation ("critical slowing down") have been longstanding obstacles to the quest for highly accurate results [5,6].

It has now clearly emerged [2] that there has been impressive progress with all these problems. A powerful strategy to deal with the problem of critical slowing down is the application of the so-called "cluster algorithms": rather than performing local Monte Carlo updates (of single degrees of freedom) in the systems, one constructs suitable non-local moves. These techniques were first introduced for the Ising model, where, by the Swendsen-Wang and Wolff cluster algorithms (named after the scientists who invented them), one constructs large connected clusters of spins having the same sign with a clever recipe. This recipe is derived such that the resulting clusters are independent of their environment and hence can be overturned at zero energy cost. These concepts have now, to some extent, been carried over to off-lattice systems in the continuum. In this way generic examples such as mixtures of hard spheres with very different sizes, a problem that hitherto was almost intractable, has become accessible to solution [2].

Also the problem of accurately locating first order phase transitions and establishing full phase diagrams of complicated model systems is now well tractable by Monte Carlo methods. Substantial progress has been achieved by "extended sampling" strategies. Many of these advanced techniques that are available today have their roots in the classical "umbrella sampling" [7,8]: the now widely applied "multicanonical Monte Carlo", "extended ensemble averaging", "multihistogram methods", etc. Some variations of these methods, such as "parallel tempering" where a number of systems are run in parallel at a set of neighboring temperatures (or other external control parameters), which are exchanged from time to time, can also be used in conjunction with Molecular Dynamics methods (the latter will be discussed below). While in some of these methods reweighting factors need to be estimated and this is not always straightforward, a particularly convenient approach is the so-called "Wang-Landau sampling" [9]: by a rather simple iteration procedure the energy density of states can be estimated, from which the free energy over the entire temperature range can then be computed. All these methods (and many related ones) are concisely described and put in perspective in Ref. [2]. Applications that are described range from the estimation of interface free energies to the energy barriers that one encounters in the folding of small proteins.

There is a number of technical problems which are more typical of classical molecular dynamics [7] than of MC, although often they are found possibly in a different shape also in MC. The first of these is how to generate averages in ensembles of statistical mechanics other than the microcanonical ensemble: in other words how to couple the system to thermostats, barostats, etc [10,11]. Another typical problem is the simultaneous occurrence of slow and fast degrees of freedom that creates difficulties for an efficient use of MD since straightforward implementations mean a great waste of computer resources. Special techniques, such as the freezing of the fast degrees of freedom by holonomic constraints [12] or multiple time step methods [13], have been developed, starting from the early times. In Ref. [2] a broad range of issues on Molecular Dynamics is spanned, from basic foundations to modern issues that are still under investigation, such as the sampling of conformational equilibria of proteins and free energy barriers.

While the topics discussed so far assume throughout that classical statistical mechanics is appropriate, it is clear that many physical phenomena of interest require a treatment based on quantum mechanics. A special case is the time evolution of a quantum particle in a classical environment. A consistent description can be based on the Wigner formulation of quantum statistical mechanics, allowing the study of transport phenomena in such mixed quantum-classical systems. Complementary approaches on related problems can be based on path integral formulations.

At this point, it should be stressed that the path integral treatment of multi-fermion problems is still hampered by a fundamental problem, the "minus sign problem". Some improvement concerning this problem is probably possible by a

resummation over paths, so the stochastic problem of sampling paths is reduced to sampling "graphs". Investigations along such lines already present in ref.[2] could well become a mainstream of research in the future.

Another new type of Quantum Monte Carlo method, the "coupled Ion-Electron Monte Carlo Method", extends the standard variational Monte Carlo approach [14] to obtain a description of electronic structure that provides an interesting alternative to the density functional based Car Parrinello [15] (CPMD) method. The advantage is that the ions can easily be treated fully quantum-mechanically, and this advantage pays off when one studies problems such as the properties of (metallic) hydrogen under high pressure.

Of course, the density functional theory and CPMD are also steadily developing further. Ref. [2] provides thorough reviews of these techniques, from their basic principles to the current state of the art. Particularly interesting extensions concentrate on the possibility to accelerate rare events and compute free energy barriers.

The sampling of rare events, *e.g.* a nucleation process where a huge free energy barrier needs to be crossed to form a critical nucleus of the new stable phase on the background of a metastable phase is a longstanding challenge to simulation. Transition path sampling has been the first scheme introduced to compute the relevant properties without assuming a reaction coordinate. In Ref. [2], both the formal theoretical background ("transition path theory") of this approach and the state of the art of the implementation of this new techniques are discussed, as well as applications to study the pathway of protein folding. (Fig. 1)

Selected applications, from simple fluids to human immune response

This article can only attempt to convey to the reader the flavour of the wide range of problems that one can address today via computer simulation in the science of condensed matter: while in many cases a reductionist approach is needed, progress is achieved, in terms of coarse-grained models, by the imaginative ideas of scientists suitably adapting new models to the problems under study. Coarse-graining means that some small-scale degrees of freedom, which are considered less relevant to the questions considered, are eliminated or else treated in simplified ways. Computer simulation offers the advantage that connections can be established between the models of condensed matter on different scales and the hierarchy from the sub-Angstrom scale, where one deals with effects due to the electrons, up to the mesoscopic and macroscopic scales relevant for living matter.

There are cases where a well-defined separation of scales occurs between relevant and irrelevant degrees of freedom (*e.g.* colloidal suspensions) and others where this is not the case (*e.g.* polymer melts, or biological membranes). In colloidal suspensions, particles of spherical or rod-like or disk-like shape with linear dimensions in the micrometer range, are immersed in a solvent fluid, which acts as a heat bath for the larger colloid particles (but presents also hydrodynamic

interactions, and, via ions or polymers dissolved in the fluid, controls also the effective interactions between the colloids). This is a classical case [2] where all the degrees of freedom of the system apart from those of the colloids can be integrated out, and models result which resemble models of simple fluids (the dynamics of the colloidal particles being, however, rather different).

The situation is very different for polymeric systems: a flexible macromolecule exhibits nontrivial structure from the Angstrom scale (chemical bonds) over the coil size (tens of nanometers) to even larger length scales of mesoscale ordering phenomena. Consequently, there does not exist a unique modelling approach to deal with matter formed from such macromolecules. Rather there exists a large diversity of proposed algorithms and different models, as outlined in depth in Ref. [2], to confront fundamental problems of polymer science, such as, for example, understanding which material properties control the "tube diameter" describing the constrained diffusion of polymers in melt (note that polymers "creep" in a snakelike motion along a "tube" that follows their own contour). (Fig 2)

Of course, there are still limitations that have not yet been overcome: while the simulations have contributed much towards the understanding of the slowing down of undercooled simple fluids near the glass transition and while one can also study glassy freezing of very short, non-entangled polymers, the interplay of chain entanglement and glassy relaxation clearly is beyond reach. Similar caveats apply to the studies of spreading and bridging of polymer droplets on substrates - there is still a huge gap between the nanometer length scale and nanosecond time scale of the simulations on the one hand, and the macroscopic scales of the experiment on the other hand. Similar caveats apply to many other problems also discussed in [2], such as defect patterns in liquid crystalline systems, surfactant layers under shear, lipid bilayers with transmembrane proteins modelling biological membranes, etc. However, one should emphasize that for problems of this type, apart from the simulations, there is no reliable theoretical information whatsoever. (Fig 3)

Even in the "classical" problem of spin glasses, a problem heavily debated and well studied for more than thirty years, there is still no relevant analytic theory apart from the mean-field theory of the infinite-range model where every spin interacts with every other spin with random Gaussian couplings.

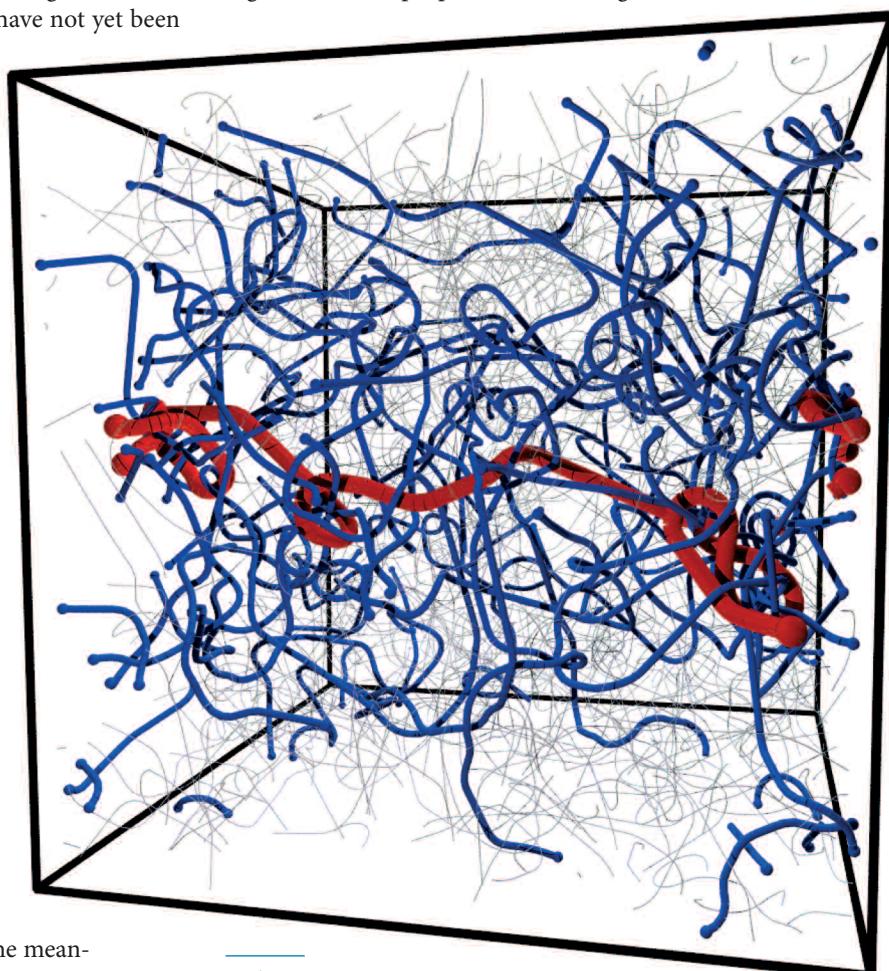
Experiments on these systems can neither probe the order parameter distribution nor the relevant correlations directly, so the guidance from simulations has been absolutely crucial for a better understanding of this problem.

Particularly challenging are the simulation tasks in biologically motivated physics: Drug-target binding necessitates quantum-mechanical approaches for the chemically reactive molecular groups, but the huge size of complex biomolecules in aqueous solution requires a simplified description on long length scales. So hybrid approaches have been introduced that combine ab initio with classical MD in the environment of the reactive groups, by suitable "quantum mechanical/molecular mechanical (QM/MM)" partitioning.

While in this problem one tries to deal with atomistic detail, the problem of evolutionary design of the so-called T-cell receptors that recognize and eliminate tumor cells requires a completely abstract model with no chemical and structural details [2].

Outlook

We have already emphasized the need to develop mixed (ab initio/classical MD) approaches, where a subsystem is treated ab-initio, and coupled to a classical environment via a transition region where a proper "handshaking" between ...



▲ Fig. 2:

Result of the "primitive path" analysis for all chains in the melt of 100 bisphenol-A polycarbonate chains of N=60 repeat units. The primitive path highlights the geometry of chain entanglement. The red chain is entangled with all those chains whose primitive paths are shown in blue, but not with those shown in light gray. Reprinted with permission from Macromolecules 38, p. 8089(205).

... both types of simulation is attempted. This is an example of "multiscale simulation", an important paradigm of the field: to combine, in a coherent fashion, complementary methods of computer simulation on vastly different time and length scales for problems of capital interest. *e.g.*, for multi-functional materials hierarchically organized from complex building blocks a molecular simulation may be impossible because of the excessive need for computer time. When one then resorts to coarse-grained models, however, there is a clear need to validate the coarse-grained model by a "mapping" to an atomistic model on smaller scales. In fact, there may also be the need to reintroduce atomistic detail consistently into coarse-grained models ("inverse mapping"). Such techniques have been proposed in the context of polymer melts, but it is clear that the problem is much more general, and far from fully solved. The goal is to develop simulation approaches for complex materials and biological matter that successfully bridge the gap from the small scales of electronic structure calculations to the mesoscopic scales of pattern formation in soft matter (where one uses coarse-grained simulation techniques such as dissipative particle dynamics, multiscale colli-

sion dynamics, etc.). While this goal will remain an exciting challenge for many years to come, nevertheless steady and important progress towards reaching this goal can be anticipated. Thus, we can expect that the role of computer simulation as the central and basic method of approach of condensed matter theory will become even much more important in the future. ■

About the authors:

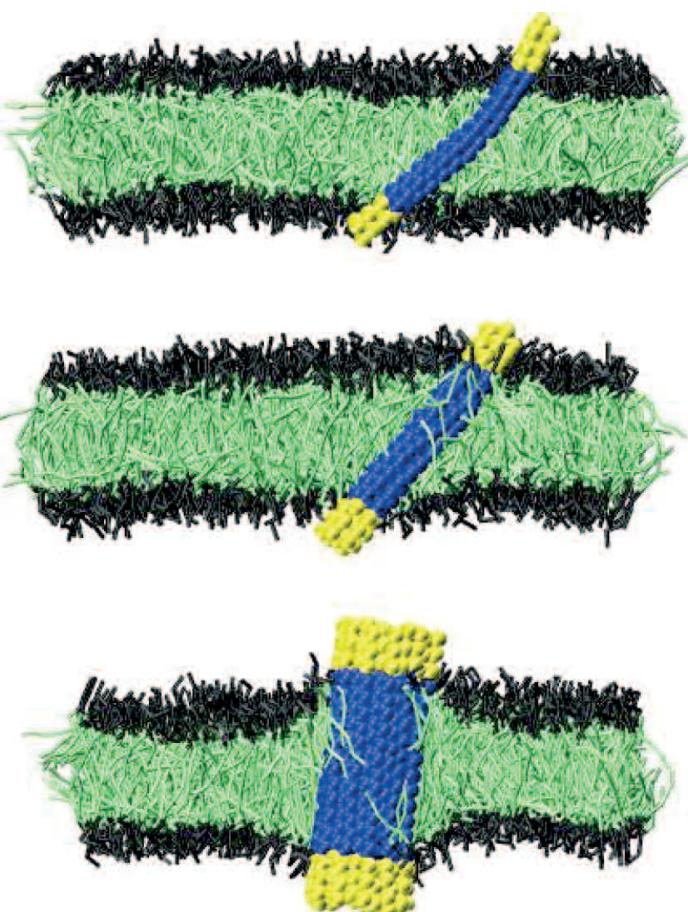
Mauro Ferrario is Professor of Structure of Matter at the University of Modena. He is a computational physicist working on interdisciplinary problems between physics and chemistry.

Giovanni Ciccotti is Professor of Structure of Matter at the University of Roma "La Sapienza". He has contributed to the development of Molecular Dynamics methods now for several decades.

Kurt Binder is Professor of Theoretical Physics at the Johannes Gutenberg Universitaat of Mainz. He belongs to the "old horses" generation in the field of Monte Carlo simulations and will receive this year the Boltzmann Medal in the field.

References

- [1] L. Yelash, M. Müller, W. Paul, and K. Binder, *PCCP* **7**, 3728-3733 (2005)
- [2] M. Ferrario, G. Ciccotti and K. Binder (eds.) *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, Vols 1,2 Springer, Berlin-Heidelberg (2006)
- [3] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Chem. Phys.* **21**, 1087 (1953)
- [4] B. J. Alder and T. E. Wainwright, in *Proceedings of the International Symposium on Statistical Mechanical Theory of Transport Processes, Brussels 1956*, ed. I. Prigogine, Interscience, NY (1958); see also, for a collection of the reprints of the papers which made the field, the book: G. Ciccotti, D. Frenkel and I. R. MacDonald (eds.) *Simulation of Liquids and Solids. MD and MC Methods in Statistical Mechanics*, North Holland, Amsterdam (1986)
- [5] K. Binder "Monte Carlo Investigations of Phase Transitions and Critical Phenomena" in C. Domb and M. S. Green (eds.) *Phase Transitions and Critical Phenomena*, Vol 5b, pp. 1-105. Academic Press, New York (1976)
- [6] D. P. Landau and K. Binder "A Guide to Monte Carlo Simulations in Statistical Physics", 2nd ed. Cambridge, Univ. Press, Cambridge (2005)
- [7] G. Ciccotti and W. G. Hoover (eds) "Molecular Dynamics Simulation of Statistical Mechanical Systems", North Holland, Amsterdam (1986)
- [8] G. M. Torrie and J. P. Valleau, *J. Comp. Phys.* **23**, 187 (1977)
- [9] F. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001)
- [10] H. C. Andersen, *J. Chem. Phys.* **72**, 2384 (1980)
- [11] S. Nosé, *Mol. Phys.* **52**, 255 (1984)
- [12] J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comp. Phys.* **23**, 327 (1977)
- [13] M. Tuckermann, B. J. Berne and G. J. Martyna, *J. Chem Phys.* **97**, 1990 (1992)
- [14] B. L. Hammond, W. A. Lester, Jr. and P. J. Reynolds "Monte Carlo methods in Ab Initio Quantum Chemistry" World Scientific, Singapore (1994)
- [15] R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985)



▲ Fig. 3: Typical peptide configurations embedded in a coarse-grained model of a biological membrane, for three different peptide sizes. Reprinted with permission from [2].